Using Gait Dynamics to Estimate Load from a Body-worn Accelerometer

James R. Williamson, Andrew Dumas, Greg Ciccarelli, Austin R. Hess, and Mark J. Buller

***Abstract*—Heavy loads increase the risk of musculoskeletal injury for foot soldiers and first responders. Continuous monitoring of load carriage in the field has proven difficult. We propose an algorithm for estimating load from a single body-worn accelerometer. The algorithm utilizes three different methods for characterizing torso movement dynamics, and maps the extracted dynamics features to load estimates using two machine learning multivariate regression techniques. The algorithm is applied to two field collections of soldiers and civilians walking with varying loads. The effect of load on features is analyzed and the feature utility in conjunction with two regression techniques is assessed. Load estimation is done using a fixed set of machine learning parameters and leave-one-subject-out cross-validation. Accurate load estimates are obtained, demonstrating robustness to changes in equipment configuration, walking conditions, and walking speeds. On soldier data with loads ranging from 45 to 89 lbs, load estimates result in mean absolute error (MAE) of 6.64 lbs and correlation of r = 0.81. On combined soldier and civilian data, with loads ranging from 0 to 89 lbs, results are MAE = 9.57 lbs and r = 0.91.**

***Index Terms*—Gait analysis, accelerometry, load estimation, regression, body sensors, walking dynamics, ambulation, correlation structure, musculoskeletal injury**

I. Introduction

Military personnel commonly engage in training and operational activities where they carry heavy loads of 35–65 kg or more, subjecting them to increased risks of musculoskeletal injury (MSI) and excessive thermal-work strain [1]. Real-time continuous monitoring of load carriage would improve the ability to assess these risks. In addition, characterizing the effect of load on gait dynamics could provide for early detection of MSI or thermal-work strain based on unusual gait patterns.

Commercially available wearable accelerometers have the potential to characterize important properties of gait beyond a laboratory setting. Our goal is practical monitoring of gait during natural walking, unencumbered by limited video field of view and allowing natural stride period variations not possible on a treadmill. A large body of work assesses gait under laboratory or clinical conditions, using stationary equipment such as treadmills, force plates, or multi-camera video systems with fiducial markers [2–4]. In contrast, using readily available sensors, we aim for a practical system for estimating load from a single torso-worn accelerometer. Challenges include non-uniform walking surfaces and speeds.

Our specific approach is to estimate load based on three types of extracted features that characterize torso movement dynamics, in conjunction with two multivariate regression techniques that map the features to load estimates. We evaluate multiple combinations of feature types and regression techniques on two field data collections, finding that the best overall performance is obtained using all three feature types and both regression techniques. This combined system results in an ability to quickly and accurately estimate load when evaluated on out-of-sample test subjects.

Distribution A: Public Release. This work was sponsored by the United States Army Research Institute of Environmental Medicine under United States Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and do not reflect the official policy of the Department of Army, Department of Defense, or the United States Government.

J. R. Williamson, A. Dumas, G. Ciccarelli, and A. R. Hess are with the Massachusetts Institute of Technology Lincoln Laboratory, Lexington MA 02421 (e-mail: jrw@ll.mit.edu).

M. J. Buller is with the United States Army Research Institute of Environmental Medicine, Natick MA 01760.

II. Methods

A. Data Collection

1) SBPE

The Soldier Protection Benchmark Evaluation (SPBE) study, consisting of 32 subjects (28 men and 4 women), was used to evaluate alternative protective equipment configurations [5]. All subjects provided informed consent and the study was authorized by both MIT and US Army institutional review boards. Acceleration data were collected at 25.6 Hz using the Equivital EQ-02 (Hidalgo Ltd., Cambridge, UK) physiological status monitor, positioned on the left side of the chest below the armpit. The Equivital EQ-02 incorporates a 3-axis ± 3 g micro electro-mechanical systems (MEMS) based accelerometer. On each day of the study, each subject wore one of four different equipment configurations, A, B, C, and D. Each configuration is associated with a different load level, due to the number of protection items it includes (Fig. 1). Data were obtained from 32 subjects for configurations A, B, and C; and from 29 subjects for configuration D.

Our analysis focused on a 5 km march that occurred in the middle of the day, generally after lunch. An automatic procedure was used to segment the 5 km march. As described in Section IIC, 60-second frames containing stable walking were automatically selected for analysis. Statistics on loads and data durations for each configuration are shown in Table I. The total march duration across all trials varies between 40 and 68 minutes. However, a smaller amount of data is automatically selected for analysis using the walking detection

algorithm described in Section IIC. The duration of analyzed data ranges between 14 and 56 minutes for all trials except for one trial that yielded only four minutes of data. The primary reason for automatic rejection of data frames was that they contained running data.

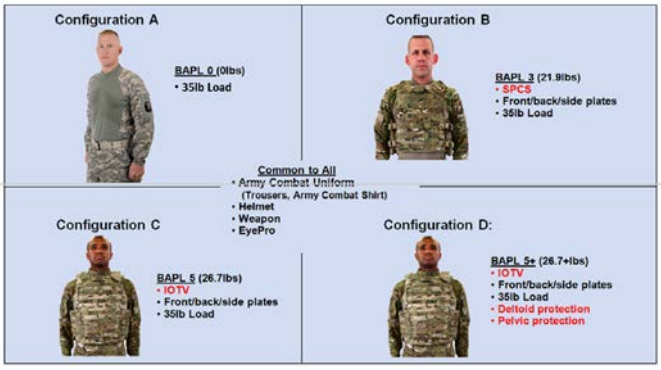

Fig. 1. Four equipment configurations worn in SPBE data collection.

TABLE I
SPBE DATA COLLECTION. STATISTICS DESCRIBING LOADS AND DURATIONS OF 5 KM MARCH DATA USED FOR ANALYSIS.

SPBE Config.	# Trials	Load (lbs)			Duration (min)		
		Min	Max	Mean	Min	Max	Mean
A	32	45.2	45.8	45.4	4	46	30
B	32	66.7	75.6	71.2	22	55	36
C	32	76.2	86.6	81.3	17	56	39
D	29	78.8	89.2	83.9	27	53	41

2) MIT-LL:

In a study approved by the MIT Committee on the Use of Humans as Experimental Subjects, a total of 31 volunteers (19 men and 12 women) between the ages of 18 and 65 wore the Equivital EQ-02 during natural walking. For 26 of these subjects, a single trial of unloaded walking data was compiled from multiple indoor and outdoor walking segments. The outdoor walking was on a looped 0.4 km gravel path that includes uneven terrain and an 8 meter elevation change. For five subjects, only outdoor walking was done, with loads of 0 lbs, 20 lbs, and (for one subject) 41 lbs. Loads were induced using a weighted backpack. Table II summarizes the statistics of the MIT-LL data collection.

TABLE II
MIT-LL DATA COLLECTION, CONSISTING OF 31 SUBJECTS. MULTIPLE LOADS WERE RECORDED FOR FIVE OF THE SUBJECTS.

Load (lbs).	# Trials	Duration (min)		
		Min	Max	Mean
0	31	12	112	55
20	5	14	24	21
41	1	20	20	20

B. Data Collection

A block diagram of the system for load estimation is shown in Fig. 2. The first processing step is feature extraction. There are three feature extraction algorithms, each of which produces summary statistics over fixed duration data frames that characterize movement dynamics. All three algorithms automatically compensate for stride duration, and therefore are insensitive to changes in walking speed. Also, because they make use of summary statistics derived over long-duration data frames, the algorithms are practical for use in real-time physiological status monitoring systems. The time duration of the data frames used in this paper is 60 seconds, with 30 second overlap between successive frames.

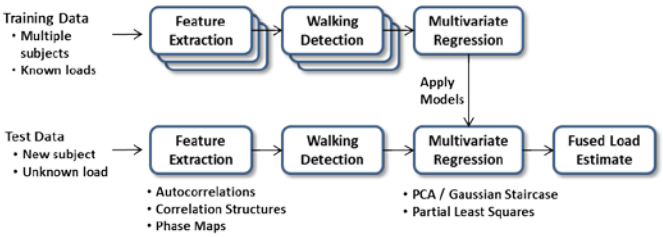

Fig. 2. Block diagram of the system for automatically estimating load from walking data. PCA = principal component analysis.

In Fig. 3 is shown acceleration data from a single SPBE subject bearing a moderate load of 45.2 lbs (left) and a heavy load of 84.2 lbs (right). These are plots of 10 s of data after conversion within a 60 s frame to standard units (z-scoring), and with offsets for easy viewing. In Fig. 4–6 are shown feature examples from the three feature types obtained from these two data frames, in order to illustrate the general feature trends that are seen between moderate and heavy loads.

1) Autocorrelation features

Stride dynamics are captured using autocorrelations of the acceleration magnitude signals. Autocorrelation peaks have been used previously to characterize gait from a single accelerometer [6]. We consider a different approach using the entire autocorrelation function out to the first peak, which represents time-delay correlation coefficients spanning a single stride. First, the autocorrelation value for accelerometer axis j in data frame i at time delay k is computed via

$$c_{j,i}(k) = \frac{1}{(n_i - k)} \sum_{t,t+k \in f_i} b_j(t) b_j(t+k), \quad (1)$$

where b_j is obtained by normalizing the accelerometer signal a_j into standard units within the 60 s data frame. Next, *time-scaled* autocorrelation patterns are formed, which span the average stride period in the data frame. This is done by selecting the peak in the longitudinal autocorrelation pattern, which is the 2nd accelerometer axis: argmax_i ($c_{2,i}$). The range of possible stride durations is constrained to be between 0.8 and 1.4 seconds, a range that easily encompasses all walking strides during this study. To obtain more precise estimates of the stride durations, cubic interpolation of the summed autocorrelation functions is used, sampled at 2.56 kHz to mitigate the effects of quantization. Time-scaling by g_i is employed at all three acceleration axes to obtain autocorrelation patterns, sampled at 49 points, k, that span one stride period:

$$d_{j,i}(k) = c_{j,i}(k/g_i). \quad (2)$$

The three time-scaled autocorrelation vectors are plotted in Fig. 4, given the data frames from Fig. 3, showing strong

differences in the Longitudinal and Lateral axes. Time-scaled autocorrelation patterns such as these have previously been used to assess gait asymmetries based on dynamics from foot-worn accelerometers [7].

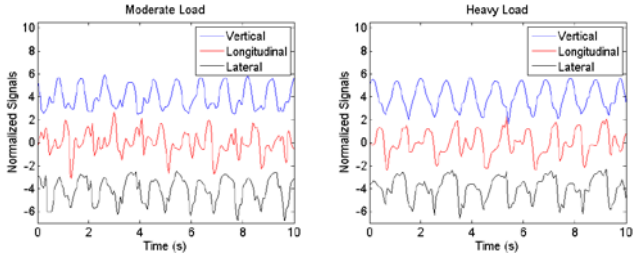

Fig. 3. Acceleration signals from the same SPBE subject with a load of 45.2 lbs (left) and 84.2 lbs (right). Signals are z-scored within 60 s frames, and offset for easy viewing.

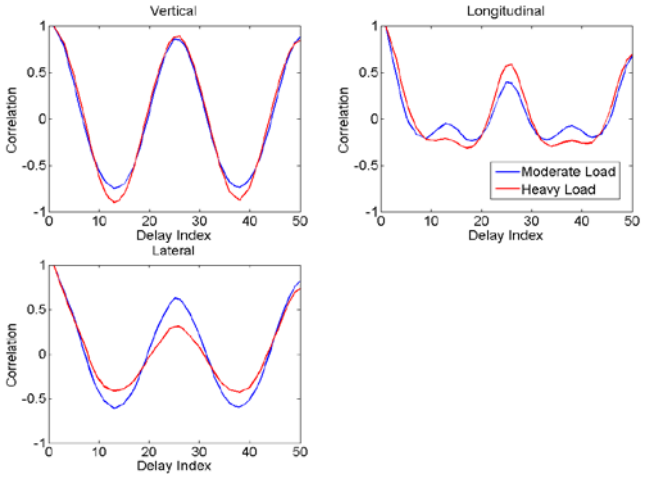

Fig. 4. Time-scaled autocorrelation patterns for the three axes, based on the data frames illustrated in Fig. 3. The Longitudinal and Lateral axes show the largest changes between the moderate and heavy loads.

2) *Correlation structure features*

The autocorrelation features described above characterize dynamics properties within each acceleration axis. Correlation structure features characterize the structure of correlation both within and across the acceleration axes. These features are insensitive to precise phase relations between channels, encoding instead a multidimensional measure of the levels of correlation. This approach is motivated by the observation that auto- and cross-correlations of measured signals can reveal hidden parameters in the stochastic-dynamical systems that generate the time series.

This multivariate feature construction approach has been applied to analysis of EEG signals for prediction of epileptic seizures [8,9], analysis of cardiorespiratory signals for prediction of infant apnea [10,11], analysis of speech signals for prediction of depression scores, [12,13], estimation of cognitive performance associated with dementia [14], and detection of changes in cognitive performance associated with mild traumatic brain injury [15]. It has also been extended to analysis of video-derived facial action unit signals for prediction of depression [11]. These previous applications all involve analysis of broadband signals, containing power across a wide range of frequencies. In this paper, the technique is applied for the first time to quasiperiodic signals in which most of the power is in a narrow frequency band that corresponds to stride frequency.

The correlation structure features are obtained from a channel-delay correlation and covariance matrix computed in each 60 s frame. Each matrix contains the product set of correlation or covariance coefficients between the three acceleration axes (channels) at 50 time delays. The delays are spaced at every 0.02 s using cubic interpolation of the acceleration signals. The 100 largest rank-ordered eigenvalues from the channel-delay correlation matrix compose the feature vector. Two additional feature elements are obtained from the channel delay covariance matrix. These are 1) the logarithm of the trace, and 2) the logarithm of the determinant of the covariance matrix. These features encode the overall power and entropy of the non-normalized acceleration signals, respectively. A detailed description of the correlation structure approach is in [9].

In Fig. 5 is illustrated the channel-delay matrices obtained from the moderate load data frame (left) and the heavy load frame (right). Below these matrices are plotted the eigenspectra for these two cases. The reduced power in lower eigenvalues (i.e., eigenvalue indices 20–100) for the heavy load reflects the general trend that heavy loading reduces dynamical complexity.

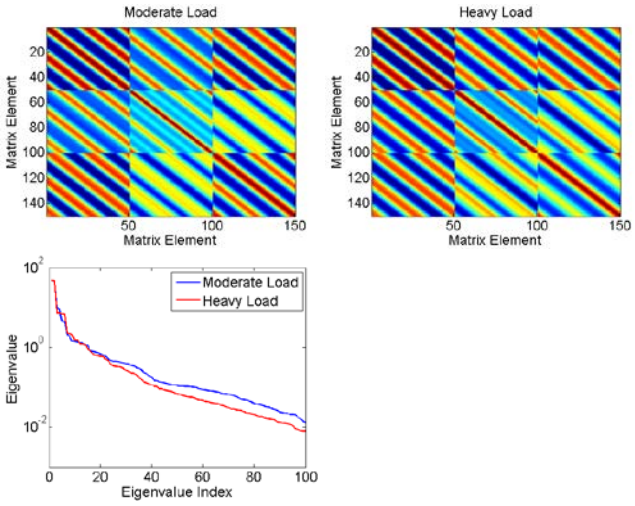

Fig. 5. Channel-delay correlation matrix for moderate load (top left) and for heavy load (top right). Eigenvalue spectrum from the two matrices shows reduced power in smaller eigenvalues for heavy load.

3) *Phase map features*

Phase map features were introduced in [1], where a 2-D (10 × 10) histogram was created in each data frame from the joint distribution of the vertical and longitudinal acceleration signals. We modify this approach by using overlapping Gaussian kernels instead of disjoint histogram bins. The Gaussian kernels provide a more effective representation, presumably due to reduced bin quantization noise. Specifically, we use a 5 × 5 array of kernels, with means centered at {-2,-1,0,1,2}, and diagonal covariance matrices with variance = 0.25. The acceleration signals were first normalized (z-scored) in each axis within each data frame.

The feature value for each of the 25 kernels is the sum, over the data frame, of its posterior probabilities given each data sample. Fig. 6 illustrates the phase map features applied to the moderate and severe load examples, which contain 1,536 data samples per frame.

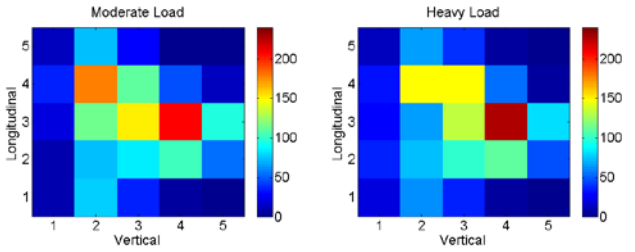

Fig. 6. Phase map features for moderate load (left) and heavy load (right), based on 60 s frames containing 1,536 data samples.

C. Walking Detection

To obtain accurate load estimates during walking, a key step is detecting frames that are composed of stable walking data. To do this, we use a procedure adapted from [7], in which k-means clustering with two clusters is first used to separate initial outliers, and then outlier frames from the dominant cluster are iteratively removed based on the distance of their time-scaled autocorrelation patterns (for the longitudinal and vertical acceleration axes) from the remaining patterns in that cluster. The walking detection algorithm is run independently in each trial, and only features from the detected, valid walking frames are used.

D. Mutivariate Regression

Our next step involves mapping the multivariate features described above into load estimates. Two methods are used for this. The first is the Gaussian Staircase (GS) method [12,13], which uses principal component analysis (PCA) to obtain lower dimensional feature vectors, then performs multivariate fusion using Gaussian mixture models (GMMs), and finally maps the GMM outputs into load estimates using univariate regression. The second is the partial least squares (PLS) technique, which directly maps the original high-dimensional feature vectors into load estimates.

1) Gaussian Staircase (GS)

GS involves three sequential processing steps. The first is dimensionality reduction of the high dimensional feature vectors obtained from the feature techniques described above. Table III lists the dimensionality reduction parameters used for the three feature types. Normalization via z-scoring is done of the correlation structure features in feature set 2 before PCA is applied. These features vary widely in magnitude, with high discriminability often found in the lower magnitude features. This motivates the need to normalize these features prior to PCA, so that all features are on the same playing field. With the feature set 1 autocorrelation features and the feature set 3 phase map features, on the other hand, normalization was found to be disadvantageous, as the natural level of variability of the features tends to covary with their discriminability.

For each feature type, the same number of principal components is used in all experiments described in Section III (Table III). The number of components was chosen empirically based on load estimation performance and is sufficient to explain a similar percentage (96-98%) of the variance of each feature set on the combined SPBE/MIT-LL data sets. During each fold of leave-one-subject-out cross-validation, the normalization and PCA coefficients are independently derived from the training set and then applied to the test data.

TABLE III
FEATURE SET DIMENSIONALITY AND PCA PARAMETERS USED BY GAUSSIAN STAIRCASE

Feature Set	# Features	z-score?	# PCA	Fraction variance
Autocorrelation	147	no	13	0.98
Corr. Struct.	102	yes	6	0.96
Phase Map	25	no	10	0.97

Next, GMMs are constructed from an ensemble of Gaussian classifiers [12,13]. The ensemble is derived by partitioning the load data into two classes, *low* and *high*, but doing this with multiple partition thresholds (5, 10, 15, …, 85 lbs). A separate Gaussian classifier is defined for each partition. Then, the ensemble of Gaussians for each class comprises the GMM for that class. For the SPBE data set, in which loads range between 45.2 and 89.2 lbs, this results in 8 Gaussians per GMM. For the combined SBPE and MIT-LL data set, in which loads range between 0.0 and 89.2 lbs, it results in 17 Gaussians per GMM. The Gaussian staircase partitioning results in a single, highly regularized GMM classifier, with feature densities that smoothly increase in the direction of decreasing load levels (Class 1) or of increasing load levels (Class 2).

A separate GMM is constructed for each feature type. The output of each GMM is the ratio of the log of the mean likelihoods, computed across all data frames in a trial, for its class. For fusion of different feature types, GMM outputs for the different feature classes are weighted according to their load estimation accuracy, which is defined by the Pearson correlation, r, of its load estimates with the true loads (see Tables IV and V). These weights are determined via [13]:

$$w_i = 1/\left(1 - r_i^2\right). \quad (3)$$

The final step is to map the GMM outputs into load estimates using a 2nd order univariate regression model, which is constructed from the training set.

2) Partial Least Squares (PLS)

PLS regression uses a projection of the independent feature variables and dependent load variables into a new space using latent factors, fitting a regression model in that new space. We use 50 latent factors and 2nd order regression. Multiple feature types are fused by concatenating the feature vectors. Also, all features are normalized (z-scored) prior to regression. PLS produces a load estimate in each data frame. These load estimates are combined across frames within the same trial using a median filter.

III. Results

A. Feature Distributions

Before evaluating the ability of the prediction algorithm to estimate load, we first analyze the distribution of features in the combined SPBE/MIT-LL data set with respect to three coarse load categories: low (0 lbs), moderate (20-46 lbs), and heavy (66-89 lbs). In making this comparison, it is important to remember that the low load is comprised entirely of data from the MIT-LL dataset, the moderate load is comprised mostly of data from the SPBE dataset, and the heavy load is comprised entirely of data from the SPBE dataset. Because the two data sets were obtained from different populations in different behavioral contexts, it is interesting to see how well the features can capture, on average, any monotonic trends due to load changes across the two disparate datasets.

Fig. 7 shows the average feature values in feature set 1 for low (blue), moderate (green), and heavy (red) loads. Specifically, the average autocorrelation feature values are plotted in the three acceleration axes. These plots can be compared to single trial results shown in Fig. 4. The notable findings are of monotonic changes for autocorrelation features in all three acceleration axes as loads progress from low to moderate to heavy. These load-based changes are smallest in the vertical axis and largest in the lateral axis.

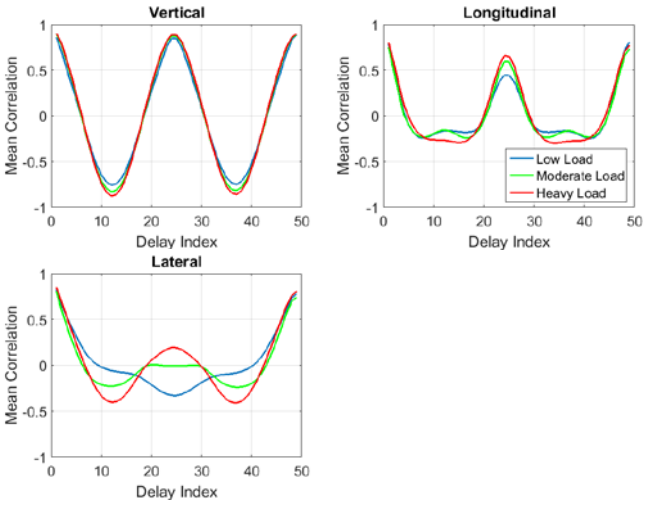

Fig. 7. Average vertical, longitudinal, and lateral autocorrelation values in feature set 1 for subjects with low loads (blue), moderate loads (green), and heavy loads (red).

Fig. 8 shows the average *normalized* feature values in feature set 2 for low, moderate, and heavy loads. Normalization via z-scoring of these features is done because the features span a wide dynamic range, requiring their normalization prior to PCA (see Table III). Due to normalization, the distance between plots indicates the Cohen's d effect size between the feature distributions. Fig. 8 (left) shows the average correlation matrix eigenvalues, ordered left to right from largest to smallest. The monotonic reduction of power in the smaller eigenvalues (indices 3-100) indicates that heavier loads produce a loss of dynamical complexity in the torso dynamics as measured by the 3-D acceleration signal.

The two other features from feature set 2, derived from the channel-delay covariance matrix, represent total entropy and power in all three acceleration axes. The entropy feature shows the same pattern as the correlation matrix eigenvalues, with a loss of entropy accompanying the progression from low to moderate to heavy loads. The power feature, on the other hand, shows a remarkably different pattern. The transition from heavy to moderate loads, consisting almost exclusively of trials from the SPBE dataset, shows an expected increase in power. However, the transition from moderate to low loads shows an unexpected decrease in power to a level below even that found for the heavy loads. This finding indicates that the power feature seems to work in a predictable way within the SPBE dataset (i.e., more power for lower load), but fails to generalize across the two datasets. The failure of this feature provides additional motivation for the use of features that characterizing acceleration dynamics based on *relative* acceleration values over time and across the three acceleration axes, as opposed to *absolute* acceleration values, which can be easily confounded across different subject populations and behavioral contexts.

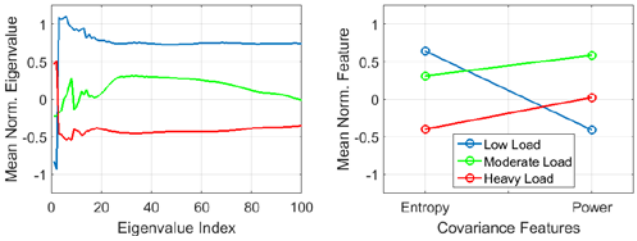

Fig. 8. Average normalized correlation structure values in feature set 2 for subjects with low loads (blue), moderate loads (green), and with heavy loads (red). Average correlation matrix eigenvalues (left), and average covariance matrix entropy and power feature (right).

Fig. 9 shows the average normalized feature values in feature set 3 for low, moderate, and heavy loads. Each set of five values corresponds to a row in the 2-D phase map (see Fig. 6). For some of these phase map features, there is a clear divergence only between low and moderate/heavy loads, whereas for others there is a clear divergence only between low/moderate and heavy loads. The lack of clear monotonic trends as seen in feature sets 1 and 2 raises questions about the ability of feature set 3 to reliably generalize across different data sets.

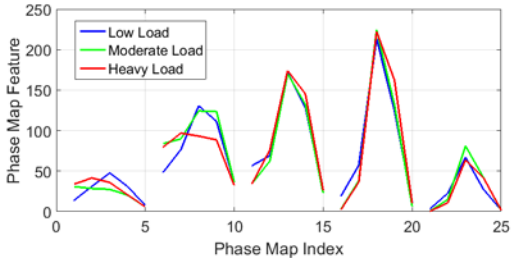

Fig. 9. Average phase map features in feature set 3 for subjects with no load (blue), moderate loads (green), and heavy loads (red). Each set of five features corresponds to a row of the 2-D phase map (see Fig. 6).

B. *Estimating Load from Field Data*

The load estimation algorithm was first applied to the SPBE data set alone. The algorithm was tested on each of the 32 subjects using training data from the remaining 31. Each feature set was evaluated separately using each regression method. In Table IV are shown the mean absolute error, MAE, and Pearson correlation, *r*, of the load estimates using estimates from GS alone, PLS alone, and the two methods fused. These results are shown for each feature set alone (rows 1–3), and for feature set combinations (rows 4–5). Fusion of the regression methods is done such that the more accurate method is weighted more highly [13], with weights based on the correlations of their load estimates when used alone,

$$w_i = r_i^2 / \left(1 - r_i^2\right). \qquad (4)$$

The best results are obtained using feature types 1,2,3 for GS and 1,2 for PLS (bottom row). In Fig. 10 (left) is plotted the estimated loads as a function of true load for the SPBE trials, using the combined system (highlighted in blue in Table IV).

TABLE IV
LOAD ESTIMATION RESULTS ON SPBE DATA SET.

Feature Sets	Gaussian Staircase		Partial Least Sq.		Combined System	
	MAE (lbs)	*r*	MAE (lbs)	*r*	MAE (lbs)	*r*
1	7.36	0.76	8.75	0.73	7.44	0.77
2	7.89	0.73	8.49	0.73	7.64	0.76
3	8.42	0.70	9.88	0.68	8.45	0.73
1,2	6.73	0.80	8.22	0.75	6.86	0.79
1,2,3	6.46	0.82	8.22	0.75	**6.64**	**0.81**

The SPBE data involves loads between 45 and 89 lbs. A natural question that arises is how well the algorithm would extend to smaller loads. To investigate this question, we augmented the SPBE data set with additional data collected from a different set of subjects at MIT-LL. Table V summarizes the results. The algorithm extends successfully to the loads ranging from 0 to 89 lbs, and shows a pattern of results similar to those obtained on SPBE alone. Compared to those results, MAE and *r* are both larger due to the larger range of loads. Fig. 10 (right) shows a scatter of true loads to estimated loads for this combined data set.

TABLE V
LOAD ESTIMATION RESULTS ON COMBINED SPBE, MIT-LL DATA SET.

Feature Sets	Gaussian Staircase		Partial Least Sq.		Combined System	
	MAE (lbs)	*r*	MAE (lbs)	*r*	MAE (lbs)	*r*
1	11.60	0.85	11.87	0.87	10.96	0.89
2	13.78	0.79	13.82	0.80	13.25	0.81
3	18.32	0.65	14.78	0.74	14.80	0.77
1,2	10.65	0.88	10.05	0.90	9.59	0.91
1,2,3	10.42	0.88	10.05	0.90	**9.57**	**0.91**

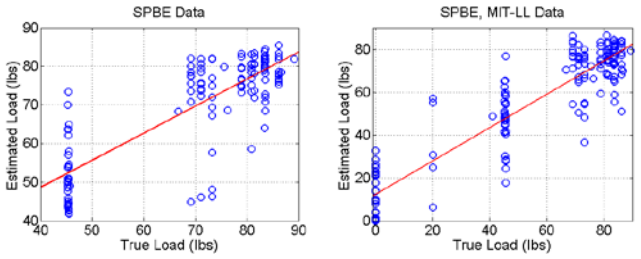

Fig. 10. Estimated load as a function of true load for the 125 SBPE trials (left) and 125 SPBE trials combined with 37 MIT-LL trials (right). Results are obtained using combined system. For SPBE, MAE = 6.64 lbs. For SPBE and MIT-LL, MAE = 9.57 lbs. Linear fits are plotted in red.

C. *Detecting Load from Field Data*

For some applications it may be more important to derive broad load classifications from the estimates [1]. To investigate the effectiveness for our load estimates for classification, we constructed receiver operating characteristic (ROC) curves. For the SPBE data set, we investigated classification of moderate (45-46 lbs) versus heavy (66-89 lbs) loads. For the combined SPBE and MIT-LL data set, we assessed classification of low (0 lbs) versus moderate/heavy (20-92 lbs) loads. The resulting ROC curves for the combined system using all three feature sets are shown in Fig. 11, with areas under the curve (AUC) of 0.96 and 0.99, respectively. The ROC curves indicate excellent sensitivity and specificity in load detection. In Table VI are shown the AUC values for the combined system given different feature type combinations.

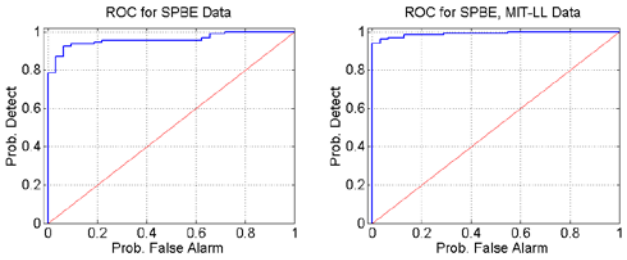

Fig. 11. Receiver operating characteristic (ROC) curves for classifying moderate versus heavy loads on SPBE data (left) and low versus moderate or heavy loads on combined SPBE, MIT-LL data (right), using three features sets and combined regression system.

TABLE VI
AREA UNDER ROC CURVES FOR COMBINED SYSTEM ON SPBE AND COMBINED SPBE, MIT-LL DATA SETS.

Feature Sets	Area Under ROC Curve (AUC)	
	SPBE Data	SPBE, MIT-LL Data
1	0.95	0.98
2	0.95	0.95
3	0.93	0.95
1,2	0.95	0.99
1,2,3	**0.96**	**0.99**

D. *Convergence of Load Estimates*

Another practical consideration is how quickly the algorithm reaches accurate load estimates. Load estimates are obtained every 30 s after the first 60 s of walking data. To investigate

the effect of data duration, we selected from each test trial 50 randomly placed contiguous data segments of varying durations. In Fig. 12 are shown MAE results from the SPBE, MIT-LL data set, given varying segment durations, for GS alone, PLS alone, and the combined system. With only one minute of data (a single data frame), the combined system achieves good performance (MAE = 11.61 lbs), which continues to improve as the data duration increases.

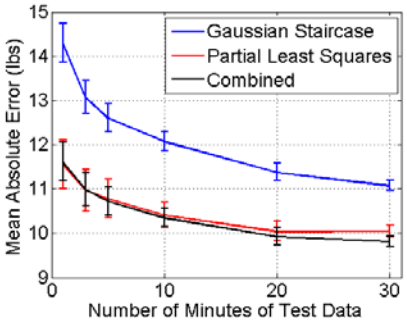

Fig. 12. MAE as a function of test data duration for each regression method and for the combined system.

E. Treadmill Study

All of the results presented thus far are obtained by training on one set of subjects and testing on a novel subject. A substantial fraction of the error may be due to inter-subject variability, which could be reduced by adapting the regression models to the test subject. Effective individualization may be possible with only unloaded training data from the test subject. To investigate individualization, we tested the algorithm on additional load data from a single MIT-LL subject for whom the field test load estimates were highly biased (estimated loads were 26.2 and 37.2 lbs given true loads of 0 and 20 lbs, respectively). We collected treadmill data at 3 mph for several different load levels. The order of loads was randomized, and two minutes of data were used for each load. Additionally, two unloaded conditions were evaluated, one without a backpack and one with an unloaded backpack.

In Fig. 13 the true loads are plotted in red, in order of load, and the estimated loads are plotted in blue. The estimates are obtained using all three feature sets and the combined system, with the regression models trained on the other subjects. The load estimates have similar biases as the previous field results, with an average bias of 23.3 lbs (28.2 lbs for loaded cases). The consistency of bias and the linear trend with increased load lend support to the idea of individualized bias removal.

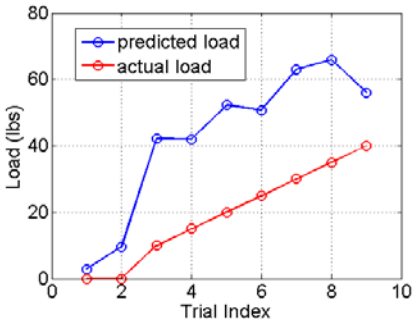

Fig. 13. Load estimates from treadmill data of single subject. Trials are two minutes duration each, and trials are ordered by load level.

IV. CONCLUSION

We demonstrate rapid, accurate estimation of load from field data collections using out-of-sample testing. The results indicate robustness to variations in body types, equipment configurations, walking conditions, and walking speeds. Future work will involve validation of the algorithms on larger and more controlled data sets, with more extensive coverage of load levels. Another exciting direction is individualization, in which feature variability due to different body types can be accounted for. We expect that the multidimensional feature techniques introduced in this paper, by characterizing complementary properties of movement dynamics, will prove useful for other gait monitoring applications, such as early detection of musculoskeletal injury, detection of excessive thermal work strain, and monitoring of physical fatigue.

Another notable finding in this paper is that the progression from low to moderate to high loads induces monotonic changes in the feature distributions for feature sets 1 and 2. The changes found in the autocorrelation patterns of feature set 1 provide clear and strong signatures of load, particularly in the lateral and longitudinal acceleration axes. The changes in eigenvalue distributions of feature set 2 provide signatures indicating a loss of multivariate complexity across the three dimensions (axes) of the acceleration signal. The same correlation structure approach has been widely used as a predictor of change in neurological state from EEG and speech (audio and video) [8-15]. Accordingly, we plan to use the gait feature techniques described in this paper to explore the effect of cognitive stress, fatigue, and neurological disorders on gait.

REFERENCES

[1] Clements, C. M., Moody, D., Potter, A. W., Seay, J. F., Fellin, R. E., & Buller, M. J. (2013, May). Loaded and unloaded foot movement differentiation using chest mounted accelerometer signatures. In *Body Sensor Networks (BSN), 2013 IEEE International Conference on* (pp. 1-5). IEEE.

[2] Patterson, K., Gage, W., Brooks, D., Black, S., & McIlroy, W. (2010).Evaluation of gait symmetry after stroke: A comparison of current methods and recommendations for standardization. Gait & Posture, 31, 241-246.

[3] Salarian, A., Russmann, H., Vingerhoets, F., Dehollain, C., Blanc, Y., Burkhard, P., & Aminian, K. (2004). Gait Assessment in Parkinsons Disease: Toward an Ambulatory System for Long-Term Monitoring. IEEE Trans. Biomedical Engineering. 51(8), 1434-1443.

[4] Kavanagh, J. & Menz, H. (2008). Accelerometry: A technique for quantifying movement patterns during walking. Gait & Posture. 28, 1-15.

[5] Hughes, T. B., Williamson, J. R., Hess, A. R., Young, W. T., Dumas, A., Fischl, K. D., & Telfer, B. A. (2013). Soldier Protection Benchmark Evaluation (SPBE) Physiological Data Collection and Analysis, Fort Greely, Alaska, 17 September-5 October 2012 (No. TR-1174). MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB.

[6] Moe-Nilssen, R., & Helbostad, J. L. (2004). Estimation of gait cycle characteristics by trunk accelerometry. Journal of biomechanics, 37(1), 121-126.

[7] Williamson, J. R., Fischl, K., Dumas, A., Hess, A., Hughes, T., & Buller, M. J. (2013, May). Individualized detection of ambulatory distress in the field using wearable sensors. In Body Sensor Networks (BSN), 2013 IEEE International Conference on (pp. 1-6). IEEE.

[8] Williamson, J. R., Bliss, D. W., & Browne, D. W. (2011, May). Epileptic seizure prediction using the spatiotemporal correlation structure of intracranial EEG. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on* (pp. 665-668). IEEE.

[9] Williamson, J. R., Bliss, D. W., Browne, D. W., & Narayanan, J. T. (2012). Seizure prediction using EEG spatiotemporal correlation structure. *Epilepsy & Behavior*, *25*(2), 230-238.

[10] Williamson, J. R., Bliss, D. W., Browne, D. W., Indic, P., Bloch-Salisbury, E., & Paydarfar, D. (2011, November). Using physiological signals to predict apnea in preterm infants. In *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on* (pp. 1098-1102). IEEE.

[11] Williamson, J. R., Bliss, D. W., Browne, D. W., Indic, P., Bloch-Salisbury, E., & Paydarfar, D. (2013, May). Individualized apnea prediction in preterm infants using cardio-respiratory and movement signals. In Body Sensor Networks (BSN), 2013 IEEE International Conference on (pp. 1-6). IEEE.

[12] Williamson, J. R., Quatieri, T. F., Helfer, B. S., Horwitz, R., Yu, B., & Mehta, D. D. (2013, October). Vocal biomarkers of depression based on motor incoordination. In *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge* (pp. 41-48). ACM.

[13] Williamson, J. R., Quatieri, T. F., Helfer, B. S., Ciccarelli, G., & Mehta, D. D. (2014, November). Vocal and Facial Biomarkers of Depression Based on Motor Incoordination and Timing. In *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge* (pp. 65-72). ACM.

[14] Yu, B., Quatieri, T. F., Williamson, J. R., & Mundt, J. C. (2014). Prediction of cognitive performance in an animal fluency task based on rate and articulatory markers. In *Fifteenth Annual Conference of the International Speech Communication Association*.

[15] Helfer, B. S., Quatieri, T. F., Williamson, J. R., Keyes, L., Evans, B., Greene, W. N., ... & Heaton, K. (2014). Articulatory Dynamics and Coordination in Classifying Cognitive Change with Preclinical mTBI. In *Fifteenth Annual Conference of the International Speech Communication Association*.

[16] Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, *185*, 1-17.